# ABSTRACT

A collection of documents, such as the documents that result from a Web search, is categorized into a hierarchy of categories based on the documents in the collection. Various types of information, such as the title of a document, a document format type, and anchor text are utilized to determine the similarity between documents and generate most relevant and accurate categories.